

(19) World Intellectual Property  
Organization  
International Bureau



534007

(43) International Publication Date  
10 June 2004 (10.06.2004)

PCT

(10) International Publication Number  
**WO 2004/049240 A1**

(51) International Patent Classification<sup>7</sup>: **G06K 9/00**  
(21) International Application Number:  
PCT/IB2002/004990

(22) International Filing Date:  
28 November 2002 (28.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **NOKIA CORPORATION** [FI/FI]; Keilalahdentie 4, FIN-02150 Espoo (FI).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **THEIMER, Wolfgang** [DE/DE]; Am Hohwege 10, 44879 Bochum (DE). **ROSS, Andree** [DE/DE]; Wittkamp 15, 44534 Lünen (DE).

(74) Agent: **KURIG, Thomas**; Becker, Kurig, Straus, Bavariastrasse 7, 80336 München (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

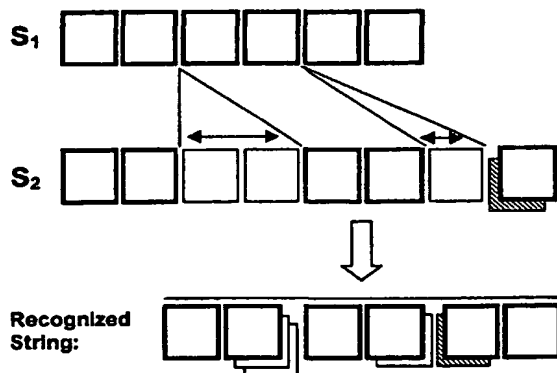
Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND DEVICE FOR DETERMINING AND OUTPUTTING THE SIMILARITY BETWEEN TWO DATA STRINGS

Similar structure of two compared strings  $S_1$  and  $S_2$ :



Legend:  
□ = Concordant Character  
■ = Missing Character  
▨ = Additional Character

(57) Abstract: The present invention discloses a method and device for determining and outputting a similarity measure between two data strings each data string comprising data entities, comprising: receiving a first data string, receiving a second data string, which is characterized by determining consecutively following data entities in said first data string, determining the relative positions of said consecutively following data entities in said first data string, determining similar data entities with the same order in said second data string, determining the relative positions of said determined data entities in said second data string, determining a matching measure by determining how far the relative positions of data entities in said second data string match with the relative positions of consecutively following data entities in said first data string, and outputting a similarity measure which corresponds to the matching measure of at least one comparison result.

WO 2004/049240 A1